

What is claimed is:

1. A computer-implemented method of indexing a database of documents, a subset of the documents containing nested fields, each nested field having an associated start meta word and end meta word, each meta word having an associated nesting level, the method comprising:

indexing each document containing nested fields by:

parsing the document to determine locations within the document of words and

meta words in the document and to determine the nesting level associated with each meta word; and

generating an index including

word entries, each word entry identifying locations within the document of an identified word;

meta word entries, each meta word entry identifying locations within the document of an identified meta word and indicating the determined nesting level associated with the meta word; and

generic meta word entries, each generic meta word entry identifying locations within the document of a class of meta words, including meta words at all nesting levels of the meta words found in the document, the generic meta word entry including, for each identified location within the generic meta word entry, information identifying the nesting level associated with the meta word at the identified location.

2. The computer-implemented method of claim 1, wherein each word entry, meta word entry, and generic meta word entry includes an object identifier and a location list.

3. The computer-implemented method of claim 1, wherein, for each word entry, the object identifier includes one or more words, and the location list includes locations of the one or more words in the document.

4. The computer-implemented method of claim 1, wherein, for each meta word entry,

the object identifier includes a meta word and an indication of the nesting level associated with the meta word, and the location list includes locations of the meta word in the document.

5. The computer-implemented method of claim 1, wherein, for each generic meta word entry, the object identifier includes a class of meta words, including meta words at all nesting levels of the meta words found in the document, and the location list includes locations of each occurrence of each meta word in the class of meta words in the document, and further includes an indication of the nesting level associated with each occurrence of each meta word in the class of meta words at each location.

6. The computer-implemented method of claim 5, wherein, for the location list for each generic meta word entry, each location of each occurrence of each meta word in the class of meta words in the document is mathematically combined with the nesting level associated with that occurrence of that meta word at that location to encode both the location and the nesting level into a single value.

7. A computer-implemented method of searching a database of documents, a subset of the documents containing nested fields, each nested field having an associated start meta word and end meta word, each meta word having an associated nesting level, the method comprising:

- receiving a query that specifies one or more words to be found within a specified field within a document;

- determining a start meta word and end meta word associated with the specified field;
- searching an index to identify locations of the specified words and locations of a class of meta words that includes at least one of the start meta word and end meta word associated with the specified field;

- applying first spatial criteria to the identified locations of the class of meta words with respect to the identified locations of the specified words to select a meta word from the class of meta words;

- determining the nesting level of the selected meta word;

- identifying a complementary meta word corresponding to the selected meta word;

16 searching the index to determine a location for the identified complementary meta
17 word; and
18 applying second spatial criteria to the identified locations of the specified words and
19 to the determined location for the identified complementary meta word to generate a result
20 that indicates whether the specified words are found within a first field associated with the
21 selected meta word and the identified complementary meta word.

1 8. The computer-implemented method of claim 7, wherein the first field is the specified
2 field.

1 9. The computer-implemented method of claim 7, wherein at least a plurality of the steps
2 of applying first spatial criteria, determining the nesting level, identifying a complementary
3 meta word, searching the index, and applying second spatial criteria are repeated until a final
4 a result is generated.

1 10. The computer-implemented method of claim 9, wherein the final result is selected
2 from the set consisting of (A) the specified words are found within the specified field, and (B)
3 there is no instance of the specified words within the specified field.

1 11. The computer-implemented method of claim 7, wherein the step of identifying
2 comprises identifying a complementary meta word corresponding to the selected meta word
3 and to its determined nesting level.

1 12. The computer-implemented method of claim 7, wherein the class of meta words
2 includes a specific meta word at all nesting levels of the specific meta word found in the
3 database.

1 13. The computer-implemented method of claim 7, wherein the step of applying first
2 spatial criteria comprises the steps of:
3 determining a closest location of the identified locations of the class of meta words
4 with respect to an identified location of the specified words, and

5 selecting the meta word from the class of meta words corresponding to the determined
6 closest location.

1 14. The computer-implemented method of claim 13, wherein the step of applying second
2 spatial criteria comprises determining whether the identified location of the specified words
3 falls between the determined location for the identified complementary meta word and the
4 determined closest location of the identified locations of the class of meta words to generate a
5 result that indicates whether the specified words are found within the specified field.

1 15. The computer-implemented method of claim 7, wherein the class of meta words
2 further includes an object identifier and a location list, the object identifier including at least
3 one of the start meta word and end meta word, and the location list including a location, and
4 nesting level information at that location, for each occurrence of the at least one of the start
5 meta word and end meta word.

1 16. A computer-implemented method of searching a database of documents, a subset of
2 the documents containing nested fields, each nested field having an associated start meta
3 word and end meta word, each meta word having an associated nesting level, the method
4 comprising:

5 receiving a query that specifies one or more words to be found within a first specified
6 field that is found within a second specified field within a document;

7 determining a first start meta word and first end meta word associated with the first
8 specified field, and a second start meta word and second end meta word associated with the
9 second specified field;

10 searching an index to identify:

11 locations of the specified words,
12 locations of a first class of meta words that includes at least one of the first
13 start meta word and first end meta word associated with the first specified field, and
14 locations of a second class of meta words that includes at least one of the
15 second start meta word and second end meta word associated with the second specified field;
16 applying first spatial criteria, determined at least in part from the received query, to
17 the identified locations of the first and second classes of meta words and the identified

locations of the specified words to select a first meta word from the first class of meta words,
and a second meta word from the second class of meta words;
determining the nesting levels of the first and second selected meta words;
identifying a first and second complementary meta words, corresponding to the first
and second selected meta words;
searching the index to determine a location for the first identified complementary
meta word and a location for the second identified complementary meta word; and
applying second spatial criteria, determined from the received query, to the identified
locations of the specified words and to the determined locations for the first and second
identified complementary meta words to generate a result that indicates whether the specified
words are found within a first field, associated with the first selected meta word and the first
identified complementary meta word, that is found within a second field, associated with the
second selected meta word and the second identified complementary meta word.

17. The computer-implemented method of claim 16, wherein the first field is the first
specified field and the second field is the second specified field.

18. The computer-implemented method of claim 16, wherein at least a plurality of the
steps of applying first spatial criteria, determining nesting levels, identifying first and second
complementary meta words, searching the index, and applying second spatial criteria are
repeated until a final a result is generated.

19. The computer-implemented method of claim 18, wherein the final result is selected
from the set consisting of (A) the specified words are found within an instance of the first
specified field that is found within an instance of the second specified field, and (B) there is
no instance of the specified words within an instance of the first specified field that is within
an instance of the second specified field.

20. The computer-implemented method of claim 16, wherein the steps of applying first
spatial criteria, determining nesting levels, identifying first and second complementary meta
words, searching the index and applying second spatial criteria include:

4 applying the first spatial criteria to the identified locations of the first class of meta
5 words and to the identified locations of the specified words to select a first meta word from
6 the first class of meta words;
7 determining the nesting level of the first selected meta word;
8 identifying a first complementary meta word, corresponding to the first selected meta
9 word;
10 searching the index to determine a location for the first identified complementary
11 meta word;
12 applying the second spatial criteria to the identified locations of the specified words
13 and to the determined location for the first identified complementary meta word to generate a
14 first result;
15 applying the first spatial criteria to the identified locations of the second class of meta
16 words and to the identified location corresponding to the first selected meta word to select a
17 second meta word from the second class of meta words;
18 determining the nesting level of the second selected meta word;
19 identifying a second complementary meta word, corresponding to the second selected
20 meta word;
21 searching the index to determine a location for the second identified complementary
22 meta word; and
23 applying the second spatial criteria to the determined location for the first identified
24 complementary meta word and to the determined location for the second identified
25 complementary meta word to generate a second result.

1 21. The computer-implemented method of claim 20, wherein at least a plurality of the
2 steps of applying first spatial criteria to the identified locations of the first class of meta
3 words, determining the nesting level of the first selected meta word, identifying the first
4 complementary meta word, searching the index to determine a location for the first identified
5 complementary meta word, and applying second spatial criteria to generate a first result are
6 repeated until the first result is a first final result.

1 22. The computer-implemented method of claim 21, wherein the first final result is

selected from the set consisting of (A) the specified words are found within an instance of the first specified field, and (B) there is no instance of the specified words within an instance of the first specified field.

23. The computer-implemented method of claim 16, wherein the step of identifying comprises identifying a first and second complementary meta words corresponding to the first and second selected meta words, and to the determined nesting levels of the first and second selected meta words.

24. The computer-implemented method of claim 16, wherein the first and second classes of meta words include a specific meta word at all nesting levels of the specific meta word found in the database.

25. The computer-implemented method of claim 16, wherein the step of applying first spatial criteria comprises the steps of:

determining a closest location of the identified locations of the first class of meta words with respect to an identified location of the specified words,

selecting a first meta word from the first class of meta words corresponding to the determined closest location of the first class of meta words,

determining a closest location of the identified locations of the second class of meta words with respect to the first closest location of the identified locations of the first class of meta words, and

selecting a second meta word from the second class of meta words corresponding to the determined closest location of the second class of meta words.

26. The computer-implemented method of claim 25, wherein the step of applying second spatial criteria comprises the steps of:

determining whether the identified location of the specified words falls between the determined location for the first identified complementary meta word and the determined closest location of the first class of meta words,

6 determining whether the determined location for the first identified complementary
7 meta word falls between the determined location for the second identified complementary
8 meta word and the identified location of the specified words, and
9 generating a result, based on the determining steps of applying second spatial criteria,
10 that indicates whether that the specified words are found within an instance of the first
11 specified field that is found within an instance of the second specified field.

1 27. The computer-implemented method of claim 16, wherein the first class of meta words
2 further includes an object identifier and a location list, the object identifier including at least
3 one of the first start meta word and first end meta word, and the location list including a
4 location, and nesting level information at that location, for each occurrence of the at least one
5 of the first start meta word and first end meta word.

1 28. The computer-implemented method of claim 16, wherein the first class of meta words
2 and the second class of meta words constitute the same class of meta words.

1 29. A computer-implemented method for searching an index of a database of documents,
2 the index having entries, each entry including an object identifier and a location list, each
3 object identifier including at least one of a word and a meta word, each location list including
4 one or more locations of the at least one of a word and a meta word of each corresponding
5 object identifier, each entry associated with a meta word also including nesting level
6 information for the meta word, the computer-implemented method comprising:

7 receiving a query that specifies one or more words to be found within a specified field
8 within a document;

9 determining a start meta word and end meta word associated with the specified field;

10 identifying a bounding meta word by selecting one of the start meta word and end
11 meta word;

12 searching the index to identify a first entry that has an object identifier associated with
13 the specified words;

14 searching the index to identify a second entry that has an object identifier
15 associated with the bounding meta word;

16 determining a bounding location from a closest occurrence of the bounding meta word
17 with respect to the specified words, by comparing the location list of the second entry and the
18 location list of the first entry;

19 identifying nesting level information for the bounding meta word at the bounding
20 location;

21 identifying a complementary meta word to the bounding meta word having
22 corresponding nesting level information as the identified nesting level information for the
23 bounding meta word;

24 searching the index to locate a third entry that has an object identifier associated with
25 the complementary meta word;

26 determining a complementary location from the location list of the third entry; and
27 generating a result that indicates whether the specified words are within a first field,
28 associated with the bounding meta word and the complementary meta word, by determining
29 whether a location in the location list of the first entry falls between the bounding location
30 and the complementary location.

31 30. The computer-implemented method of claim 29, wherein the complementary location
32 of the complementary meta word is opposite to the bounding location of the bounding meta
33 word, relative to the specified words, in the index of the database of documents.

1 31. The computer-implemented method of claim 29, wherein the first field is the specified
2 field.

1 32. The computer-implemented method of claim 29, wherein the step of determining a
2 bounding location comprises determining a bounding location of the bounding meta word by
3 applying first spatial criteria to the location list of the second entry and the location list of the
4 first entry, and further wherein the computer-implemented method includes repeating at least
5 a plurality of the steps of determining a bounding location, identifying nesting level
6 information, identifying a complementary meta word, searching the index, determining a
7 complementary location, and generating a result until a final a result is generated.

1 33. The computer-implemented method of claim 32, wherein the final result is selected

2 from the set consisting of (A) the specified words are found within the specified field, and (B)
3 there is no instance of the specified words within the specified field.

1 34. A computer program product for use in conjunction with a computer system, the
2 computer system for indexing a database of documents, a subset of the documents containing
3 nested fields, each nested field having an associated start meta word and end meta word, each
4 meta word having an associated nesting level, the computer program product comprising a
5 computer readable storage medium and a computer program mechanism embedded therein,
6 the computer program mechanism comprising:

7 an indexer for indexing each document containing nested fields by configuring the
8 computer to:

9 parse the document to determine locations within the document of words and
10 meta words in the document and to determine the nesting level associated with each meta
11 word; and

12 generate an index including

13 word entries, each word entry identifying locations within the
14 document of an identified word;

15 meta word entries, each meta word entry identifying locations
16 within the document of an identified meta word and indicating the determined nesting level
17 associated with the meta word; and

18 generic meta word entries, each generic meta word entry
19 identifying locations within the document of a class of meta words, including meta words at
20 all nesting levels of the meta words found in the document, the generic meta word entry
21 including, for each identified location within the generic meta word entry, information
22 identifying the nesting level associated with the meta word at the identified location.

1 35. The computer program product of claim 34, wherein each word entry, meta word
2 entry, and generic meta word entry includes an object identifier and a location list.

1 36. The computer program product of claim 34, wherein, for each meta word entry, the
2 object identifier includes a meta word and an indication of the nesting level associated with
3 the meta word, and the location list includes locations of the meta word in the document.

1 37. The computer program product of claim 34, wherein, for each generic meta word
2 entry, the object identifier includes a class of meta words, including meta words at all nesting
3 levels of the meta words found in the document, and the location list includes locations of
4 each occurrence of each meta word in the class of meta words in the document, and further
5 includes an indication of the nesting level associated with each occurrence of each meta word
6 in the class of meta words at each location.

1 38. The computer program product of claim 37, wherein, for the location list for each
2 generic meta word entry, each location of each occurrence of each meta word in the class of
3 meta words in the document is mathematically combined with the nesting level associated
4 with that occurrence of that meta word at that location to encode both the location and the
5 nesting level into a single value.

1 39. A computer program product for use in conjunction with a computer system, the
2 computer system for searching a database of documents, a subset of the documents containing
3 nested fields, each nested field having an associated start meta word and end meta word, each
4 meta word having an associated nesting level, the computer program product comprising a
5 computer readable storage medium and a computer program mechanism embedded therein,
6 the computer program mechanism comprising:

7 instructions for receiving a query that specifies one or more words to be found within
8 a specified field within a document;

9 instructions for determining a start meta word and end meta word associated with the
10 specified field;

11 instructions for searching an index to identify locations of the specified words and
12 locations of a class of meta words that includes at least one of the start meta word and end
13 meta word associated with the specified field;

14 instructions for applying first spatial criteria to the identified locations of the class of
15 meta words with respect to the identified locations of the specified words to select a meta
16 word from the class of meta words;

17 instructions for determining the nesting level of the selected meta word;

18 instructions for identifying a complementary meta word corresponding to the selected
19 meta word;
20 instructions for searching the index to determine a location for the identified
21 complementary meta word; and
22 instructions for applying second spatial criteria to the identified locations of the
23 specified words and to the determined location for the identified complementary meta word
24 to generate a result that indicates whether the specified words are found within a first field
25 associated with the selected meta word and the identified complementary meta word.

1 40. The computer program product of claim 39, wherein the first field is the specified
2 field.

1 41. The computer program product of claim 39, wherein at least a plurality of the
2 instructions for applying first spatial criteria, determining the nesting level, identifying a
3 complementary meta word, searching the index, and applying second spatial criteria are
4 repeated until a final a result is generated.

1 42. The computer program product of claim 41, wherein the final result is selected from
2 the set consisting of (A) the specified words are found within the specified field, and (B)
3 there is no instance of the specified words within the specified field.

1 43. The computer program product of claim 39, wherein the instructions for identifying
2 comprise instructions for identifying a complementary meta word corresponding to the
3 selected meta word and to its determined nesting level.

1 44. The computer program product of claim 39, wherein the class of meta words includes
2 a specific meta word at all nesting levels of the specific meta word found in the database.

1 45. The computer program product of claim 39, wherein the instructions for applying first
2 spatial criteria comprise:
3 instructions for determining a closest location of the identified locations of the class
4 of meta words with respect to an identified location of the specified words, and

instructions for selecting the meta word from the class of meta words corresponding to the determined closest location.

46. The computer program product of claim 45, wherein the instructions for applying second spatial criteria comprise instructions for determining whether the identified location of the specified words falls between the determined location for the identified complementary meta word and the determined closest location of the identified locations of the class of meta words to generate a result that indicates whether the specified words are found within the specified field.

47. The computer program product of claim 39, wherein the class of meta words further includes an object identifier and a location list, the object identifier including at least one of the start meta word and end meta word, and the location list including a location, and nesting level information at that location, for each occurrence of the at least one of the start meta word and end meta word.

48. A computer program product for use in conjunction with a computer system, the computer system for searching a database of documents, a subset of the documents containing nested fields, each nested field having an associated start meta word and end meta word, each meta word having an associated nesting level, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for receiving a query that specifies one or more words to be found within a first specified field that is found within a second specified field within a document;

instructions for determining a first start meta word and first end meta word associated with the first specified field, and a second start meta word and second end meta word associated with the second specified field;

instructions for searching an index to identify:

locations of the specified words,

locations of a first class of meta words that includes at least one of the first start meta word and first end meta word associated with the first specified field, and

locations of a second class of meta words that includes at least one of the second start meta word and second end meta word associated with the second specified field; instructions for applying first spatial criteria, determined at least in part from the received query, to the identified locations of the first and second classes of meta words and the identified locations of the specified words to select a first meta word from the first class of meta words, a second meta word from the second class of meta words; instructions for determining the nesting levels of the first and second selected meta words; instructions for identifying a first and second complementary meta words, corresponding to the first and second selected meta words, and searching the index to determine a location for the first identified complementary meta word and a location for the second identified complementary meta word; and instructions for applying second spatial criteria, determined from the received query, to the identified locations of the specified words and to the determined locations for the first and second identified complementary meta words to generate a result that indicates whether the specified words are found within a first field, associated with the first selected meta word and the first identified complementary meta word, that is found within a second field, associated with the second selected meta word and the second identified complementary meta word.

49. The computer program product of claim 48, wherein the first field is the first specified field and the second field is the second specified field.

50. The computer program product of claim 48, wherein at least a plurality of the instructions for applying first spatial criteria, determining nesting levels, identifying first and second complementary meta words, searching the index, and applying second spatial criteria are repeated until a final result is generated.

51. The computer program product of claim 50, wherein the final result is selected from the set consisting of (A) the specified words are found within an instance of the first specified field that is found within an instance of the second specified field, and (B) there is no instance

of the specified words within an instance of the first specified field that is within an instance of the second specified field.

52. The computer program product of claim 48, wherein the instructions for applying first spatial criteria, determining nesting levels, identifying first and second complementary meta words, searching the index and applying second spatial criteria include:

instructions for applying the first spatial criteria to the identified locations of the first class of meta words and to the identified locations of the specified words to select a first meta word from the first class of meta words;

instructions for determining the nesting level of the first selected meta word;

instructions for identifying a first complementary meta word, corresponding to the first selected meta word;

instructions for searching the index to determine a location for the first identified complementary meta word;

instructions for applying the second spatial criteria to the identified locations of the specified words and to the determined location for the first identified complementary meta word to generate a first result;

instructions for applying the first spatial criteria to the identified locations of the second class of meta words and to the identified location corresponding to the first selected meta word to select a second meta word from the second class of meta words;

instructions for determining the nesting level of the second selected meta word;

instructions for identifying a second complementary meta word, corresponding to the second selected meta word;

instructions for searching the index to determine a location for the second identified complementary meta word; and

instructions for applying the second spatial criteria to the determined location for the first identified complementary meta word and to the determined location for the second identified complementary meta word to generate a second result.

53. The computer program product of claim 52, wherein at least a plurality of the instructions for applying first spatial criteria to the identified locations of the first class of meta words, determining the nesting level of the first selected meta word, identifying the first

complementary meta word, searching the index to determine a location for the first identified complementary meta word, and applying second spatial criteria to generate a first result are repeated until the first result is a first final result.

54. The computer program product of claim 53, wherein the first final result is selected from the set consisting of (A) the specified words are found within an instance of the first specified field, and (B) there is no instance of the specified words within an instance of the first specified field.

55. The computer program product of claim 48, wherein the instructions for identifying comprise instructions for identifying a first and second complementary meta words corresponding to the first and second selected meta words, and to the determined nesting levels of the first and second selected meta words.

56. The computer program product of claim 48, wherein the first and second classes of meta words include a specific meta word at all nesting levels of the specific meta word found in the database.

57. The computer program product of claim 48, wherein the instructions for applying first spatial criteria comprise:

- instructions for determining a closest location of the identified locations of the first class of meta words with respect to an identified location of the specified words,
- instructions for selecting a first meta word from the first class of meta words corresponding to the determined closest location of the first class of meta words,
- instructions for determining a closest location of the identified locations of the second class of meta words with respect to the first closest location of the identified locations of the first class of meta words, and
- instructions for selecting a second meta word from the second class of meta words corresponding to the determined closest location of the second class of meta words.

58. The computer program product of claim 51, wherein the instructions for applying second spatial criteria comprise:

instructions for determining whether the identified location of the specified words falls between the determined location for the first identified complementary meta word and the determined closest location of the first class of meta words,

instructions for determining whether the determined location for the first identified complementary meta word falls between the determined location for the second identified complementary meta word and the identified location of the specified words, and

instructions for generating a result that indicates whether that the specified words are found within an instance of the first specified field that is found within an instance of the second specified field if both determinations are true.

59. The computer program product of claim 48, wherein the first class of meta words further includes an object identifier and a location list, the object identifier including at least one of the first start meta word and first end meta word, and the location list including a location, and nesting level information at that location, for each occurrence of the at least one of the first start meta word and first end meta word.

60. The computer program product of claim 48, wherein the first class of meta words and the second class of meta words constitute the same class of meta words.

61. A computer program product for use in conjunction with a computer system, the computer system for searching an index of a database of documents, the index having entries, each entry including an object identifier and a location list, each object identifier including at least one of a word and a meta word, each location list including one or more locations of the at least one of a word and a meta word of each corresponding object identifier, each entry associated with a meta word also including nesting level information for the meta word, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for receiving a query that specifies one or more words to be found within a specified field within a document;

instructions for determining a start meta word and end meta word associated with the specified field;

13 instructions for identifying a bounding meta word by selecting one of the start meta
14 word and end meta word;
15 instructions for searching the index to identify a first entry that has an object identifier
16 associated with the specified words;
17 instructions for searching the index to identify a second entry that has an object
18 identifier associated with the bounding meta word;
19 instructions for determining a bounding location from a closest occurrence of the
20 bounding meta word with respect to the specified words, by comparing the location list of the
21 second entry and the location list of the first entry;
22 instructions for identifying nesting level information for the bounding meta word at
23 the bounding location;
24 instructions for identifying a complementary meta word to the bounding meta word
25 having corresponding nesting level information as the identified nesting level information for
26 the bounding meta word;
27 instructions for searching the index to locate a third entry that has an object identifier
28 associated with the complementary meta word;
29 instructions for determining a complementary location from the location list of the
30 third entry; and
31 instructions for generating a result that indicates whether the specified words are
32 within a first field, associated with the bounding meta word and the complementary meta
33 word, by determining whether a location in the location list of the first entry falls between the
34 bounding location and the complementary location.

1 62. The computer program product of claim 61, wherein the complementary location of
2 the complementary meta word is opposite to the bounding location of the bounding meta
3 word, relative to the specified words, in the index of the database of documents.

1 63. The computer program product of claim 61, wherein the first field is the specified
2 field.

1 64. The computer program product of claim 61, wherein the instructions for determining a

2 bounding location comprise instructions for determining a bounding location of the bounding
3 meta word by applying first spatial criteria to the location list of the second entry and the
4 location list of the first entry, and further wherein the computer program product includes
5 instructions for repeating at least a plurality of the instructions for determining a bounding
6 location, identifying nesting level information, identifying a complementary meta word,
7 searching the index, determining a complementary location, and generating a result until a
8 final a result is generated.

1 65. The computer program product of claim 64, wherein the final result is selected from
2 the set consisting of (A) the specified words are found within the specified field, and (B)
3 there is no instance of the specified words within the specified field.